

Black Holes and the Holographic Principle

Lárus Thorlacius

Science Institute, Dunhaga 3, 107 Reykjavik, Iceland
 lth@hi.is

Abstract

This lecture reviews the black hole information paradox and briefly appraises some proposed resolutions in view of developments in string theory. It goes on to give an elementary introduction to the holographic principle.

1 Introduction

The theory of black holes involves a subtle interplay between gravity and quantum physics. Semiclassical arguments indicate that the time evolution of a system, where a black hole forms and then evaporates, cannot be governed by the standard postulates of quantum mechanics. If the black hole forms by gravitational collapse from an initial matter configuration that is nonsingular and described by a pure quantum state, and if Hawking radiation is truly thermal, then the formation and evaporation process evolves a pure state into a mixed one, in violation of quantum mechanical unitarity. Alternatively, unitarity may be maintained in black hole evolution but at the price of giving up locality at a fundamental level.

The long-standing debate regarding these issues, initiated by Hawking [1, 2], played a key role in the development of the *holographic principle* [3, 4]. This radical principle goes beyond black hole physics. It concerns the number of degrees of freedom in nature and states that the entropy of matter systems is drastically reduced compared to conventional quantum field theory. This claim is supported by the covariant entropy bound [5] which is valid in a rather general class of spacetime geometries. The notion of holography is well developed in certain models and backgrounds, in particular in the context of the adS/cft correspondence. A more general formulation is lacking, however, and the ultimate role of the holographic principle in fundamental physics remains to be identified.

2 Black hole evolution

Let us start by reviewing the basic ingredients that go into the black hole information paradox. We consider a black hole formed in gravitational collapse and let us assume that the initial matter configuration is approximately spherical and sufficiently diffuse so that spacetime curvature is everywhere small at early times. The subsequent evolution, including the formation and evaporation of the black hole, is then well represented by

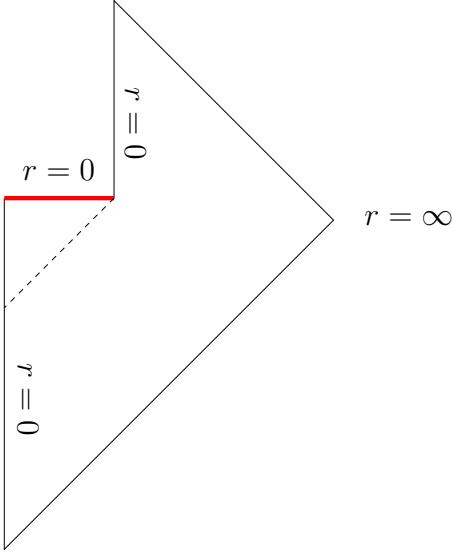


Figure 1: Penrose diagram for a semiclassical black hole geometry. At early and late times the geometry approaches that of Minkowski spacetime.

the Penrose diagram in Figure 1. The diagram assumes exact spherical symmetry. Only radial and time coordinates are displayed, with each point in the diagram representing a transverse two-sphere, whose area depends on the radial coordinate.

Penrose diagrams are a useful means of portraying the global geometry and causal properties of a given spacetime. Without going into details we note two key properties. First of all, the coordinates are chosen in such a way that radially directed light-rays correspond to straight lines oriented at ± 45 degrees to the vertical axis. The second feature is that, following a conformal mapping that brings infinity to a finite point, the entire spacetime geometry is represented by a finite region. As a result, the causal relationship between any two events is easily read off a Penrose diagram but proper distances in spacetime are not faithfully depicted.

The event horizon, shown as a dotted line in Figure 1, defines the boundary of the black hole region, inside which timelike observers cannot avoid running into the future singularity. An important feature of black holes is that local gravitational effects are extremely weak at the event horizon of a large black hole. In fact, any curvature invariant formed from the Riemann tensor will go as an inverse power of the black hole mass. For a Schwarzschild black hole, for example, one finds

$$R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} = \frac{3}{4l_{\text{Pl}}^2} \frac{M_{\text{Pl}}^4}{M^4} \quad (1)$$

at the event horizon. This observation forms the basis of the semiclassical approach to black holes, which assumes that only low-energy physics is involved in the formation and evaporation of a black hole, except in the region near the singularity, and that away from this region physics can be described by a local effective field theory. The detailed

construction of such an effective field theory for black hole evolution presents a formidable and unsolved technical problem, but let us for the moment assume that such a theory can be found and sketch the argument for information loss in black hole evolution.

2.1 Semiclassical information loss

The first step is to choose appropriate spatial slices through the black hole spacetime to provide a set of Cauchy surfaces for the quantum evolution of our system. The initial slice is taken to lie in the asymptotic past where spacetime is approximately flat and contains a diffuse distribution of matter that will later undergo gravitational collapse. The final slice, at asymptotically late times, contains a long train of outgoing Hawking radiation and possibly also a Planck mass remnant of the black hole.¹ It turns out to be possible to choose spatial slices at an intermediate stage in such a way that they lie partly inside the black hole region and partly outside, as indicated in Figure 2. For a large enough black hole this can be done in such a way that the following two requirements are met:

1. There are Cauchy surfaces that intersect the worldlines of the infalling matter inside the black hole but also the worldlines of most of the outgoing Hawking radiation that is emitted during the black hole lifetime.
2. These spatial slices avoid the region of strong curvature near the black hole singularity and are also smooth in the sense that their extrinsic curvature is everywhere small.

An explicit construction of a family of *nice slices* of this type is for example given in [6].

The semiclassical theory of black hole evolution rests on the assumption that, given such a family of Cauchy surfaces, the dynamics of the combined matter and gravity system is governed by a low-energy effective field theory and that no Planck scale effects enter into the physics except near the curvature singularity. The argument for information loss is based on the existence of this effective field theory but not on its detailed form.

Suppose an initial configuration of collapsing matter in a weakly curved background is described by a pure quantum state $|\psi(\Sigma_{\text{in}})\rangle$ defined on the surface Σ_{in} in Figure 2. The Hamiltonian of the effective field theory generates a linear evolution of this state into another pure state $|\psi(\Sigma_{\text{P}})\rangle$ on the surface Σ_{P} , which is partially inside and partially outside the black hole region. The inside and outside portions of Σ_{P} , denoted by Σ_{bh} and Σ_{ext} respectively, are spacelike separated and as a result all observables in the effective field theory that have support on Σ_{ext} commute with observables that have support on Σ_{bh} . The state on Σ_{P} is therefore an element of a tensor product Hilbert space,

$$|\psi(\Sigma_{\text{P}})\rangle \in \mathcal{H}_{\text{bh}} \otimes \mathcal{H}_{\text{ext}}. \quad (2)$$

¹The final stage of the evaporation process is governed by Planck scale physics, which we have limited knowledge of, and thus we cannot preclude the existence of black hole remnants. The important question for the information problem is not whether remnants exist but rather how many distinct remnant states are possible. We discuss black hole remnants in Section 2.2.2.

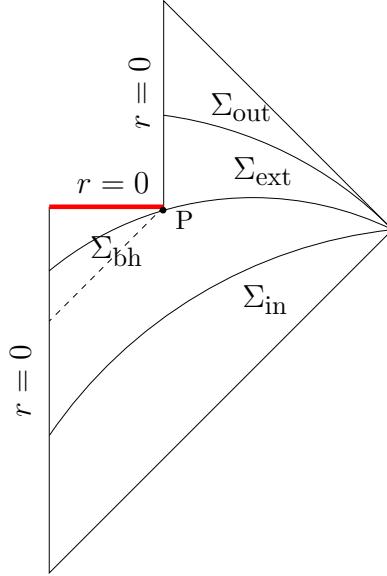


Figure 2: Cauchy surfaces used in the argument for information loss in black hole evolution.

Observers outside the black hole have no access to the part of $|\psi(\Sigma_P)\rangle$ that is inside the black hole and ultimately runs into the singularity. As a result we are instructed to trace over states in \mathcal{H}_{bh} giving rise to a mixed state density matrix on \mathcal{H}_{ext} . This mixed state will then evolve into another mixed state on the late time Cauchy surface Σ_{out} .

We now have a paradox on our hands, for if the entire process of black hole formation and evaporation is to preserve unitarity, then the final configuration of the system must be described by a pure quantum state $|\psi(\Sigma_{\text{out}})\rangle$. This final state is obtained from the initial state by a unitary S-matrix,

$$|\psi(\Sigma_{\text{out}})\rangle = S |\psi(\Sigma_{\text{in}})\rangle, \quad (3)$$

where $S S^\dagger = \mathbf{1}$. This relation can in principle be inverted to express the initial state in terms of the final one,

$$|\psi(\Sigma_{\text{in}})\rangle = S^\dagger |\psi(\Sigma_{\text{out}})\rangle, \quad (4)$$

In other words, in a unitary theory the final quantum state carries all information that is contained in the initial state. The semiclassical argument, on the other hand, resulted in a mixed final state from which there is no way to recover the initial state. In other words, quantum information is lost in semiclassical black hole evolution.

2.2 Proposed resolutions

There are a number of ways to respond to this paradox. Let us briefly review the three main proposals.

2.2.1 Information loss

Hawking advocated that the above semiclassical argument should be taken at face value and that an added fundamental uncertainty is to be incorporated into quantum physics when gravitational effects are taken into account [2]. He also had a concrete proposal involving a modified set of axioms for quantum field theory that allows pure states to evolve into mixed ones. In Hawking's formalism the unitary S-matrix of conventional quantum field theory, which maps an initial quantum state to a final quantum state, is replaced by a superscattering operator $\$,$ which maps an initial density matrix to a final density matrix, and processes involving black hole formation, or even virtual processes involving gravitational fluctuations, give rise to superscattering that mixes quantum states.

This proposal was criticised by a number of authors [7, 8]. In particular, Banks *et al.* argued that Hawking's density matrix formalism is equivalent to conventional quantum field theory coupled to randomly fluctuating sources everywhere in spacetime. As a result the theory does not have empty vacuum as its ground state but rather a thermal configuration at the Planck temperature [7]. This is clearly phenomenologically unacceptable but it should be noted that the argument rests on certain technical assumptions and conceivably a loophole may be found to avoid the thermal disaster. The early criticism appears to have put a stop to further developments in this direction and the general view is that Hawking's density matrix formalism is not a viable option for resolving the information paradox. This, of course, does not rule out a theory incorporating information loss being developed in the future.

At the moment, however, our best candidate for a theory of quantum gravity is superstring theory and this theory does not favor information loss. In its original formulation, string theory is an S-matrix theory and as such it is manifestly unitary. On the other hand, the original formulation of string theory really only amounts to a perturbative prescription for scattering amplitudes and is not adequate for describing macroscopic processes such as the formation of a large mass black hole. On the other hand, we now have non-perturbative formulations of string theory both in an asymptotically flat background [9] and in anti-de Sitter spacetime [10], where the gravitational dynamics has a dual description which is unitary. Admittedly, in both cases the duality is founded on conjectures that are unproven, and will be difficult to prove in all generality because they involve weak-strong coupling dualities, but supporting evidence has been pouring in for several years now, building a strong case. We will return to this in section 4.

2.2.2 Black hole remnants

An alternative viewpoint, put forward by Aharonov *et al.* [11], is that black holes do not completely evaporate but rather leave behind remnants that are stable or extremely long-lived. Quantum mechanical unitarity is then maintained by having the black hole remnant carry information about the initial quantum state of infalling matter that forms the black hole.

If we assume that Hawking's semiclassical calculation of particle emission remains valid

until the remaining black hole mass approaches the Planck scale and that none of the initial information goes out with the Hawking radiation then the mass of a black hole remnant can be no more than a few times the Planck mass and there needs to be a distinct remnant for every possible initial state. As a result, the density of these remnant states at the Planck energy must be virtually infinite and this leads to severe phenomenological problems if the remnants behave at all like localized objects. Their effect on low-energy physics could then be described in terms of an effective field theory and contributions from virtual remnant states would dominate almost any quantum process one might consider. Even if the amplitude for producing any given Planck mass remnant as an intermediate in, say, e^+e^- scattering at a collider is extremely small, the infinite density of such states would nevertheless make remnants the dominant channel. An infinite density of states also leads to a divergent pair production rate of remnants in weak background fields and to divergent thermal sums. Since these effects are not observed either the information carried by a black hole is not left behind in a Planck scale remnant or such remnants are described by very unconventional laws of physics at low energy.

About ten years ago some remnant models were suggested, where these pathologies were to be avoided by accommodating the high density of states in a large internal volume carried by the remnant and connected to the rest of spacetime via a Planck scale throat region [12]. Although the models had some success and served as a warning against drawing too firm a conclusion from arguments based on effective field theory, they have not been developed further. A major reason for this can be traced to subsequent developments in string theory. With the advent of string duality and branes the basic degrees of freedom of string theory have more or less been identified, and they do not include exotic black hole remnants at the Planck scale.

One of the triumphs of string theory in the nineties was the microphysical explanation and direct calculation of the Bekenstein-Hawking entropy of certain extremal and near-extremal black holes [13]. The entropy is obtained by counting configurations of strings and branes that carry the same charges as the black hole in question. Such counting is only reliable for a weakly coupled collection of strings and branes in flat spacetime which bears little resemblance to the strongly curved geometry of a black hole, but for a configuration that corresponds to an extremal black hole one appeals to extended supersymmetry and non-renormalization theorems to argue that the counting will, in fact, also hold at strong coupling where the system is more appropriately described as a black hole. The notion that, by moving around inside the parameter space of the theory, one can find a dual description of black holes in terms of weakly coupled (highly excited) strings and branes [14] leaves no room for a divergent density of remnant states at the Planck energy.

2.2.3 Black hole complementarity

A third possibility, pioneered by Page [15] and 't Hooft [16], is that Hawking radiation is not exactly thermal but in fact carries all the information about the initial state of the infalling matter. This information must then be encoded in subtle correlations between quanta emitted at different times during the evaporation process because even if the formation and

evaporation process as a whole is governed by a unitary S-matrix the radiation emitted at any given moment will appear thermal. Detecting the information would require statistical analysis of a large number of observations made on an ensemble of black holes formed from identically prepared initial states. This is a conservative viewpoint in that it assumes unitarity in all quantum processes, also when gravitational effects are taken into account, but it leads to a novel view of spacetime physics and requires us to give up the notion of locality at a fundamental level.

The question is whether the infalling matter will give up all information about its quantum state to the outgoing Hawking radiation or whether the information gets carried into the black hole. If the information is imprinted on the Hawking radiation then it must also be removed from the infalling matter for otherwise we would have a duplication of the information in a quantum state in violation of the linearity of quantum evolution. We can compare this to the more conventional information loss when a book is burned. All the information that was originally contained in the book can in principle be learned from measurements on the outgoing smoke and radiation, but at the same time this information is no longer to be found in the remains of the book. In this case, however, it is a well understood microphysical process that removes the information from the book and transfers it to the outgoing radiation, whereas matter in free fall entering a large black hole does not encounter any disaster before passing through the event horizon.

The principle of *black hole complementarity* [17] states that there is no contradiction between outside observers finding information encoded in Hawking radiation, and having observers in free fall pass through the event horizon unharmed. The validity of this principle requires matter to have unusual kinematic properties at very high energy but it does not conflict with known low-energy physics. Contradictions only arise when we attempt to directly compare the physical description in widely different reference frames. The laws of nature are the same in each frame and low-energy observers in any single frame cannot establish duplication of information [18].

2.2.4 The stretched horizon

In order to illustrate the concept of black hole complementarity it is useful to have a physical picture of the evaporation process in the outside frame. For some time astrophysicists have made use of the membrane paradigm of black holes to describe the classical physics of a quasistationary black hole [19]. From the point of view of outside observers the black hole is then replaced by a *stretched horizon*, which is a membrane placed near the event horizon and endowed with certain mechanical, electrical and thermal properties. This description is dissipative and irreversible in time. One does not have to be specific about the location of the stretched horizon as long as it is close to the event horizon compared to the typical length scale of the astrophysical problem under study.

In the context of black hole evaporation one goes a step further and views the classical stretched horizon as a coarse grained thermodynamic description of an underlying microphysical system, a quantum stretched horizon, located a Planck distance outside the event horizon and with a number of states given by $\exp(A/4)$, where A is the black hole area in

Planck units [17]. The nature of the microphysics involved was left unspecified in [17] but it was later suggested that the dynamics of the stretched horizon might be explained in the context of string theory [20]. The sticky part is that, in order to implement black hole complementarity, we have to stipulate that this membrane, or stretched horizon is absent in the reference frame of an observer entering the black hole in free fall.

The evaporation of a large black hole is a slow process and, for the purpose of our discussion here, the evolving geometry may be approximated by a static Schwarzschild solution,

$$ds^2 = -(1 - \frac{2M}{r})dt^2 + (1 - \frac{2M}{r})^{-1}dr^2 + r^2d\Omega^2. \quad (5)$$

Provided the mass M is sufficiently large, an observer in free fall will not experience any discomfort upon crossing the event horizon at $r = 2M$ but a so-called fiducial observer, who is at rest with respect to the Schwarzschild coordinate frame, will be bathed in thermal radiation at a temperature that depends on the radial position,

$$T(r) = \frac{1}{8\pi M}(1 - \frac{2M}{r})^{-1/2}. \quad (6)$$

This temperature diverges near the black hole, $T \approx (2\pi\delta)^{-1}$, where δ is the proper distance between the fiducial observer and the event horizon. The radiation can be attributed to the acceleration required to remain stationary at fixed r , which diverges as $\delta \rightarrow 0$. In the spirit of the membrane paradigm the radiation can also be viewed as thermal radiation emanating from a hot stretched horizon. The region nearest the event horizon, where the temperature (6) formally diverges, is then replaced by a Planck temperature membrane located a proper distance of one Planck length outside the event horizon.

The stretched horizon is then the source of Hawking radiation as far as outside observers are concerned. It emits Planck energy particles but most of these will fall back into the black hole. Those who escape its gravitational pull predominantly carry low angular momentum and are redshifted to energies of order the Hawking temperature by the time they reach the asymptotic region. In this view, Hawking radiation is the enormously redshifted glow from a Planck scale inferno at the stretched horizon. It carries all information about the initial state of the matter that formed the black hole, albeit in a severely scrambled form.

No infalling observer survives the encounter with the hot membrane. In fact nothing ever enters the black hole in the reference frame of outside observers, including the matter that formed the black hole in the first place. It is familiar from classical general relativity that matter falling into a black hole appears to slow down as it approaches the horizon and fades out of view due to the gravitational redshift. This is after all where the term black hole comes from. In the quantum membrane picture, the infalling matter runs into the stretched horizon, gets absorbed into it and thermalized, and is then slowly returned back out along with the rest of the black hole as it evaporates.

This scenario conflicts with the notion that infalling observers feel no ill effects upon passing through the event horizon of a large black hole. The analysis of several gedanken experiments, designed to expose possible contradictions between the experience of infalling observers and the description in the outside reference frame, led to the conclusion that the

apparent contradictions could in each case be traced to assumptions about physics at or above the Planck scale [18]. This observation does not by itself resolve the information problem but it challenges the underlying assumption of the semiclassical approach that the paradox can be posed in terms of low-energy physics alone without any reference to the Planck scale.

If we take the principle of black hole complementarity at face value we have to accept a radically new view of spacetime physics. The notion of a local event, that is invariant under coordinate transformations, is central to general relativity. According to black hole complementarity observers in different reference frames can totally disagree about the location of the rather significant event where an observer falling into a black hole meets his or her end. The proper distance between the event horizon and the final singularity is proportional to the black hole mass and can therefore be arbitrarily large. The principle therefore introduces a new degree of relativity into fundamental physics beyond the familiar relativity of measuring sticks and time pieces. It requires physics to be non-local on arbitrarily large lengthscales, yet conventional locality and causality must be recovered in everyday processes at low energy.

A detailed physical theory of black hole evaporation that incorporates black hole complementarity has not been developed.² There are indications, however, from string theory that such a description should be possible, especially in the context of the adS/cft correspondence.³ In this case, the dual gauge theory is unitary and its S-matrix in principle includes processes where black holes are formed in the adS background and subsequently evaporate [22]. The problem is that the spacetime interpretation of gauge theory observables is obscure and it is non-trivial to establish local causality at low energy in adS spacetime in the language of the gauge theory, even in the absence of black holes [23].

3 The holographic principle

Local quantum field theory leads to unitarity violation in the context of black hole evolution. This can be avoided by adopting the viewpoint of black hole complementarity but then spacetime physics is required to be non-local at a fundamental level. This notion of non-locality was taken further by 't Hooft [3] and Susskind [4], who argued that the number of available quantum states in any given region is much smaller than one might naively expect. The claim is that this number is not an extensive quantity, *i.e.* one that scales as the volume of the region in question, as one would find in any local quantum field theory with an ultraviolet cutoff, but is instead proportional to a surface area associated with the region. This dramatic reduction in the number of states as compared to conventional quantum field theory is referred to as the holographic principle.

This principle only arises when gravitational effects are taken into account. Let us

²After this lecture was delivered, Horowitz and Maldacena have proposed to reconcile Hawking's semiclassical arguments with a unitary S-matrix by imposing a final state boundary condition at the black hole singularity [21]. Their proposal may provide a realization of black hole complementarity.

³We briefly describe the holographic nature of this correspondence in Section 4.

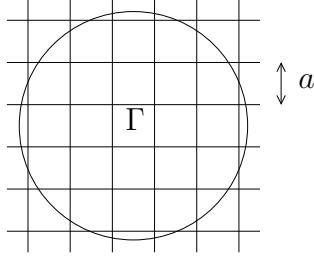


Figure 3: Spherical region Γ inside a cubic lattice.

illustrate this by a simple example. Neglect gravity for the moment and consider a three-dimensional cubic lattice of lattice spacing a and with a spin located at each lattice site. Let Γ be a spherical region as indicated in Figure 3. Let V and A be the volume and area of Γ and assume that V is large compared to a^3 . Then the total number of spins inside Γ is given by V/a^3 and if each spin is a two-level system the total number of states available to the spins in Γ is $2^{V/a^3}$. The maximal entropy of the spin system in Γ is given by the logarithm of the number of available states,

$$S_{\max} = \frac{\log 2}{a^3} V. \quad (7)$$

As expected, the maximal entropy is proportional to V . In fact, any local quantum field theory, regularized by an ultraviolet cutoff such as the lattice in our example, gives rise to a maximal entropy that is proportional to the spatial volume of the system.

This result turns out to be quite wrong when gravity is included. To see this, we carry out a thought experiment sometimes referred to as the Susskind process. The system Γ is placed at the center of a large imploding shell of matter that carries energy just such that a black hole of area A is formed when the shell collapses into itself and engulfs the spins in Γ . We now compare the maximal entropy of the system, that consists of Γ along with the collapsing shell, to that of the resulting black hole,

$$S_{\text{initial}} = S_{\Gamma} + S_{\text{shell}}, \quad (8)$$

$$S_{\text{final}} = S_{\text{BH}} = \frac{1}{4} A. \quad (9)$$

By Bekenstein's generalized second law of thermodynamics the black hole entropy must be the greater of the two and as a result the maximal entropy of the spin system in Γ is bounded by $1/4$ of its area in Planck units,

$$S_{\text{final}} \geq S_{\text{initial}} \Rightarrow S_{\Gamma} \leq \frac{1}{4} A, \quad (10)$$

instead of being proportional to the volume V . Why does quantum field theory so grossly overestimate the maximal entropy? The answer has to do with gravitational back reaction.

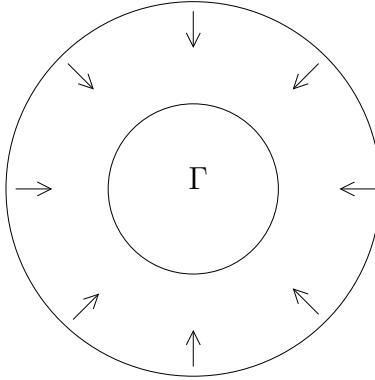


Figure 4: The Susskind process: The region Γ is placed inside an imploding shell that forms a black hole with area A .

Most of the states of the spin system are highly excited and will curve the surrounding spacetime to the extent that it collapses to a large black hole, in fact much larger than the one formed in the Susskind process.

3.1 The spacelike entropy bound

A useful way to state the holographic nature of spacetime physics is through entropy bounds that relate the number of available states to a surface area associated with a given quantum system rather than its volume.⁴ We start with a simple entropy bound motivated by the analysis of the Susskind process.

- Spacelike entropy bound: *The entropy contained in any spatial region will not exceed 1/4 of the area of the region's boundary in Planck units.*

This bound turns out to be too naive and is only valid under very restrictive assumptions. It is nevertheless useful as a step towards the much more universal covariant entropy bound of Bousso [5] which we state later on. It is instructive to consider some of the objections to the spacelike entropy bound.

(i) *Particle species:* The entropy of a matter system depends on the number of species of particles in the theory, and the entropy bound will be violated if this number is sufficiently large. Just how large the number has to be depends on the size of the system. A simple estimate [24] shows that the spacelike entropy bound fails for a weakly coupled gas in a volume with surface area A in Planck units if the number of massless particle species is $N > A$. For a surface area of 1 cm^2 the required number of species is enormous, $N \sim 10^{66}$. As there is no upper bound on the possible number of species in a field theory the entropy bound can formally always be violated. On the other hand, it is too much to expect an arbitrary field theory coupled to gravity to give sensible results. The holographic principle

⁴For a detailed review of the holographic principle and various entropy bounds see [24].

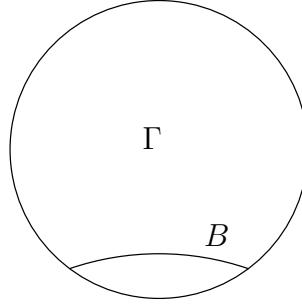


Figure 5: The region Γ almost fills the closed universe.

is put forward as a law of nature and the number of light matter fields in the real world is relatively small.

Note also that we arrived at the spacelike entropy bound by analysing the Susskind process, which involves the formation of a black hole, but this black hole is violently unstable if N exceeds the black hole area in Planck units, due to the large number of available channels for Hawking radiation [25]. We would therefore not expect the spacelike entropy bound to hold in such a theory.

(ii) *Closed FRW universe*: Consider a Friedmann-Robertson-Walker cosmological solution with positive spatial curvature. Spatial slices at equal cosmic time have the geometry of a three-sphere. Take Γ to be a large comoving region centered around an observer at the 'north pole' of the three-sphere, as indicated in Figure 5. Now let Γ be so large that its boundary $B(\Gamma)$ approaches the south pole. In this limit the region Γ consists of the entire spatial universe while the area of its boundary shrinks to zero and the spacelike entropy bound is clearly violated.

(iii) *Spatially flat FRW universe*: Finally we consider another cosmological setup, where Γ is taken to be a spherical region of proper radius R in a spatially flat FRW spacetime. The proper volume and surface area of this region are given by

$$V(\Gamma) = \frac{4\pi}{3} R^3 \quad A(\Gamma) = 4\pi R^2. \quad (11)$$

Let us assume that this universe is filled with an isotropic and homogeneous background radiation carrying a uniform entropy density σ . Then the total matter entropy contained in Γ is

$$S_{\text{matter}}(\Gamma) = \sigma V(\Gamma) = \frac{4\pi}{3} \sigma R^3. \quad (12)$$

For any non-vanishing entropy density the spacelike entropy bound is violated when R is sufficiently large,

$$R > \frac{3}{4\sigma} \Rightarrow S_{\text{matter}}(\Gamma) > \frac{1}{4} A(\Gamma). \quad (13)$$

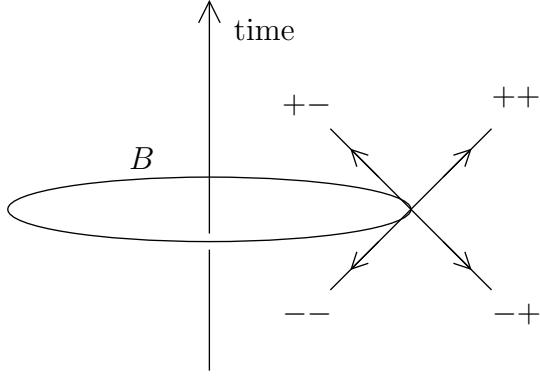


Figure 6: The circle represents a spherical surface B (with the polar angle suppressed). The lines denote future and past directed lightrays orthogonal to B .

3.2 Lightsheets and the covariant entropy bound

The foregoing examples demonstrate the failure of the spacelike entropy bound but this does not mean that the holographic principle fails. What is needed is a more geometric entropy bound that adapts to dynamical situations like the ones in examples (ii) and (iii). This is provided by Bousso's covariant entropy bound [5], which involves light-cones rather than spacelike volume, but before we get to that we need to introduce a few geometric concepts. We do this in the context of a simple example to avoid making the discussion too technical.

Consider a spherical surface B , with area $A(B)$, at rest in flat spacetime and imagine emitting light simultaneously from the entire surface at some time t_0 . The light front will propagate in two directions, radially inwards and radially outwards. One can also consider light arriving radially at B at time t_0 from the outside and inside respectively. Thus there are four families of lightrays orthogonal to B ,

- $++$ future directed, outgoing,
- $+-$ future directed, ingoing,
- $-+$ past directed, outgoing,
- $--$ past directed, ingoing.

There is nothing special about our spherical surface in this respect. Every surface in Lorentzian geometry has four orthogonal light-like directions, two future directed and two past directed, and the corresponding families of lightrays trace out null hypersurfaces (at least locally) in spacetime.

The *expansion* θ of a family of lightrays, that are orthogonal to a surface, is positive (negative) if the rays are diverging (converging) as one moves along them away from the surface. In our simple example the sign of the expansion is easily determined. The location at infinitesimal time $t = \Delta$ (or $t = -\Delta$ for past directed lightrays) as measured in the rest frame of B , of photons that were at B at time $t = 0$ defines a new surface B' and we are interested in the area of B' relative to B . If $A(B') > A(B)$ the expansion of this particular family of lightrays is positive. Beyond our example, the expansion of a family of lightrays

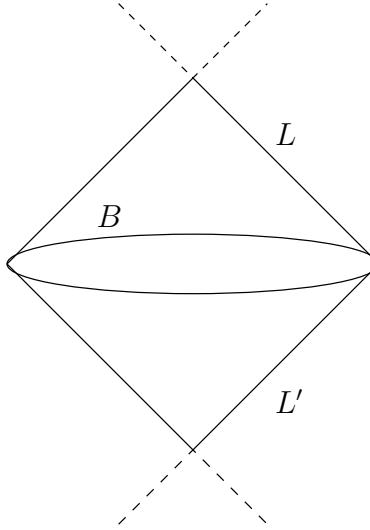


Figure 7: A lightsheet L of a surface B is a lightlike hypersurface traced out by following a family of converging lightrays orthogonal to B .

orthogonal to any smooth surface in curved spacetime can be defined locally in a coordinate invariant manner [26] and its sign determined by comparing areas of neighboring surfaces intersected orthogonally by the lightrays.

The surface B in our example is in a *normal* region where outgoing lightrays (both future and past directed) have positive expansion and ingoing lightrays have negative expansion,

$$\theta_{++} > 0, \quad \theta_{+-} < 0, \quad \theta_{-+} > 0, \quad \theta_{--} < 0. \quad (14)$$

In a *future trapped* region on the other hand both the future directed families of lightrays have negative expansion,

$$\theta_{++} < 0, \quad \theta_{+-} < 0, \quad \theta_{-+} > 0, \quad \theta_{--} > 0. \quad (15)$$

Such behavior is for example found in the collapsing region inside a black hole. There are other possibilities besides normal and future trapped but one finds in all cases that at least two out of the four families of orthogonal lightrays have non-positive expansion, $\theta \leq 0$, locally at the surface. In degenerate cases this can be true of three or even all four families.

A *lightsheet* L of a surface B is a lightlike hypersurface obtained by following a family of lightrays that is orthogonal to B and has $\theta \leq 0$. It follows from our previous comment that every surface has at least two lightsheets. The expansion will in general change as we move along L and the lightsheet ends where θ becomes positive. For example, when converging lightrays self-intersect their expansion turns positive and they no longer form a lightsheet. In other words, lightsheets do not extend beyond focal points as indicated in Figure 7 for our spherically symmetric example.

- Covariant entropy bound: *The entropy on any lightsheet of a surface B will not exceed $1/4$ of the area of B in Planck units.*

The covariant entropy bound appears to be universally applicable. At least there are no physically relevant counterexamples known. It has even been proven in the context of general relativity with the added assumption that entropy can be described by a continuum fluid and with some plausible conditions relating entropy density and energy density [27]. Of course, as was pointed out in [27], entropy at a fundamental level is not a fluid and the assumed conditions relating entropy and energy are not always satisfied. It should therefore be stressed that the covariant entropy bound has not been derived from first principles, after all the first principles of quantum gravity are unknown, but is rather an observation about the nature of matter and gravity that should be explained by a fundamental theory.

Under certain assumptions the covariant entropy bound implies the spacelike bound of Section 3.1 [5] but the covariant bound is valid much more generally. It is instructive to see how the covariant bound deals with the various objections that were offered to the spacelike bound.

(i) *Particle species*: The species problem is the same as in the spacelike case and is equally relevant (or irrelevant) here.

(ii) *Closed FRW universe*: Lightrays directed towards the north pole of the three sphere from B in Figure 5, *i.e.* ones that traverse Γ , have positive expansion. Therefore they do not trace out a lightsheet and the covariant entropy bound does not apply. Both lightsheets of B are directed towards the south pole in this case and the covariant entropy bound is valid for the complement of Γ .

(iii) *Spatially flat FRW universe*: Without going into details we note that the problem here had to do with very large regions in a spatially flat universe. The surface of such a region will have a past directed lightsheet but this lightsheet terminates at the initial singularity and has therefore much less entropy on it than is contained in the original spatial volume. It can be shown that the entropy that passes through the lightsheet is less than a quarter of the surface area of the region under consideration in Planck units [28].

4 The adS/cft correspondence

The holographic principle is put forward as a basic principle of physics and as such it should be manifest in any successful fundamental theory. It is therefore natural to ask to what extent superstring theory, the leading candidate for a unified quantum theory of matter and gravity, is holographic. Since string theory is far from a finished product, with major conceptual problems unsolved, it may be premature to subject it to this test. Yet, remarkably, it has already produced a setting where holography is explicitly realized. We finish this lecture with a quick sketch of the adS/cft correspondence and an order-of-magnitude estimate showing how the number of degrees of freedom in this nonperturbative definition of string theory is in line with the holographic principle.

The setting for the original adS/cft correspondence [10] is the physics of N coincident Dirichlet three-branes in ten-dimensional spacetime. This physics is governed by an action of the form

$$S = S_{\text{bulk}} + S_{\text{brane}} + S_{\text{int}}. \quad (16)$$

Here S_{bulk} is the ten-dimensional gravitational action, *i.e.* type IIB supergravity along with α' corrections from string theory, while S_{brane} is the worldvolume action of the N Dirichlet-branes, *i.e.* $d = 4$, $\mathcal{N} = 4$ supersymmetric $\text{SU}(N)$ Yang-Mills theory along with its own α' corrections. S_{int} describes the coupling between the ten-dimensional bulk and the branes and this coupling can be ignored in the limit of weak string coupling, $Ng_s \ll 1$.

A stack of coincident Dirichlet three-branes has a dual description as an extended object in supergravity. The corresponding ten-dimensional metric is given by the line element

$$ds^2 = \frac{1}{\sqrt{H(r)}} \left(-dt^2 + \sum_{i=1}^3 dx_i dx_i \right) + \sqrt{H(r)} \left(dr^2 + r^2 d\Omega_5^2 \right), \quad (17)$$

where $H(r) = 1 + \frac{R^4}{r^4}$. The parameter R is a characteristic length which must be large compared to the string length $\sqrt{\alpha'}$ in order for the supergravity solution to be valid. The parameters of the two descriptions are related through

$$R^4 = 4\pi g_s \alpha'^2 N, \quad (18)$$

so the supergravity description requires $g_s \ll 1 \ll Ng_s$.

The three-brane geometry has an event horizon at $r = 0$ in these coordinates, and due to gravitational redshift the near horizon region $r \ll R$ is effectively decoupled from the bulk ten-dimensional supergravity. The metric of the near-horizon region is given by

$$ds^2 \simeq R^2 \left[z^2 \left(-d\tilde{t}^2 + \sum_{i=1}^3 d\tilde{x}_i d\tilde{x}_i \right) + \frac{dz^2}{z^2} + d\Omega_5^2 \right], \quad (19)$$

where we have introduced dimensionless variables through $r = Rz$, $t = R\tilde{t}$ and $x_i = R\tilde{x}_i$. This is the metric of $\text{adS}_5 \times \text{S}_5$ using Poincaré coordinates for the adS_5 part.

The adS/cft correspondence follows from the observation that the above two descriptions of the D3-brane system both involve two decoupled factors, and in each case one of the factors is bulk ten-dimensional supergravity. By identifying the other factors with each other we are led to a duality between the worldvolume $\text{SU}(N)$ gauge theory, which is a four-dimensional conformal field theory, and string theory in the $\text{adS}_5 \times \text{S}_5$ near-horizon geometry. The two dual descriptions apply at different coupling strength, $Ng_s \ll 1$ *vs.* $g_s \ll 1 \ll Ng_s$.

4.1 AdS/cft and the holographic principle

We now give an argument, due to Susskind and Witten [29], that the number of degrees of freedom in adS -gravity in fact satisfies a holographic bound. For this it is more convenient to adopt so-called cavity coordinates for adS_5 ,

$$ds^2 = R^2 \left[- \left(\frac{1+u^2}{1-u^2} \right) d\tau^2 + \frac{4}{(1-u^2)^2} \left(du^2 + u^2 d\Omega_3^2 \right) + d\Omega_5^2 \right]. \quad (20)$$

The total entropy of the gravitational system is infinite because the spatial proper volume of adS spacetime diverges. In cavity coordinates the spatial boundary of the adS geometry is located at $u = 1$ and we impose a cutoff at $u = 1 - \epsilon$ with $\epsilon \ll 1$. With this infrared regulator in place the spatial volume is finite. The proper area of the spatial boundary is also finite and can easily be obtained from (20). This leads to the following holographic entropy bound,

$$S \leq \frac{1}{4} \times \text{“area”} \sim \left(\frac{R}{\epsilon}\right)^3 R^5 = \frac{R^8}{\epsilon^3}, \quad (21)$$

up to factors of order one.

The dual field theory is scale invariant and so it also has infinite entropy. This time around the divergence is an ultraviolet effect and can be regulated by introducing a short-distance cutoff. An important feature of the adS/cft correspondence is that infrared effects in adS space appear in the ultraviolet in the dual gauge theory. We therefore take the short distance cutoff in the field theory to be proportional to ϵ . $SU(N)$ Yang-Mills theory has $O(N^2)$ “gluon” fields so the total entropy in the dual gauge theory then goes like

$$S \sim \frac{N^2}{\epsilon^3}, \quad (22)$$

which is seen to saturate the holographic bound (21) when we take into account the relation (18) between R and N .

5 Discussion

We reviewed the black hole information problem and argued that developments in string theory strongly favor its resolution in terms of unitary evolution. This comes at a price of introducing a fundamental nonlocality into physics, but historically this nonlocality in black hole evolution served to motivate the holographic principle, a far reaching new paradigm for quantum gravity.

Many interesting aspects of holography were not touched upon here. The goal was to convey some of the basic ideas rather than give a survey of the field, which is by now quite wide. Recent work has included a revival of interest in matrix models of two-dimensional gravity [30], which provide a relatively simple realization of the holographic principle, and also the application of holographic ideas to cosmology [31].

This work was supported in part by grants from the Research Fund of the Icelandic Science and Technology Policy Council and the University of Iceland Research Fund.

References

- [1] S. W. Hawking, *Commun. Math. Phys.* **43** (1975) 199–220.
- [2] S. W. Hawking, *Phys. Rev.* **D14** (1976) 2460; *Commun. Math. Phys.* **87** (1982) 395.

- [3] G. 't Hooft, [gr-qc/9310026](#).
- [4] L. Susskind, *J. Math. Phys.* **36** (1995) 6377–6396, [hep-th/9409089](#).
- [5] R. Bousso, *JHEP* **06** (1999) 028, [hep-th/9906022](#); *JHEP* **07** (1999) 004, [hep-th/9905177](#).
- [6] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, *Phys. Rev. D* **52** (1995) 6997–7010, [hep-th/9506138](#).
- [7] T. Banks, L. Susskind, and M. E. Peskin, *Nucl. Phys.* **B244** (1984) 125.
- [8] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, and M. Srednicki, *Nucl. Phys.* **B241** (1984) 381; D. J. Gross, *Nucl. Phys.* **B236** (1984) 349; A. Strominger, [hep-th/9410187](#).
- [9] T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, *Phys. Rev. D* **55** (1997) 5112–5128, [hep-th/9610043](#).
- [10] J. M. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231–252, [hep-th/9711200](#); S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, *Phys. Lett.* **B428** (1998) 105–114, [hep-th/9802109](#); E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253–291, [hep-th/9802150](#).
- [11] Y. Aharonov, A. Casher, and S. Nussinov, *Phys. Lett.* **B191** (1987) 51.
- [12] T. Banks and M. O'Loughlin, *Phys. Rev. D* **47** (1993) 540–553, [hep-th/9206055](#); T. Banks, M. O'Loughlin, and A. Strominger, *Phys. Rev. D* **47** (1993) 4476–4482, [hep-th/9211030](#); A. Strominger and S. P. Trivedi, *Phys. Rev. D* **48** (1993) 5778–5783, [hep-th/9302080](#); S. B. Giddings, *Phys. Rev. D* **49** (1994) 947–957, [hep-th/9304027](#); *Phys. Rev. D* **49** (1994) 4078–4088, [hep-th/9310101](#).
- [13] A. Strominger and C. Vafa, *Phys. Lett.* **B379** (1996) 99–104, [hep-th/9601029](#).
- [14] G. T. Horowitz and J. Polchinski, *Phys. Rev. D* **55** (1997) 6189–6197, [hep-th/9612146](#).
- [15] D. N. Page, *Phys. Lett.* **B95** (1980) 244.
- [16] G. 't Hooft, *Nucl. Phys.* **B335** (1990) 138–154; *Phys. Scripta* **T36** (1991) 247–252.
- [17] L. Susskind, L. Thorlacius, and J. Uglum, *Phys. Rev. D* **48** (1993) 3743–3761, [hep-th/9306069](#).
- [18] L. Susskind and L. Thorlacius, *Phys. Rev. D* **49** (1994) 966–974, [hep-th/9308100](#).
- [19] K. S. Thorne, R. H. Price, and D. A. Macdonald (eds.), “Black holes: The membrane paradigm,” New Haven, USA: Yale Univ. Pr. (1986) 367p.

- [20] L. Susskind, *Phys. Rev. Lett.* **71** (1993) 2367–2368, [hep-th/9307168](#); *Phys. Rev.* **D49** (1994) 6606–6611, [hep-th/9308139](#).
- [21] G. T. Horowitz and J. Maldacena, *JHEP* **02** (2004) 008, [hep-th/0310281](#).
- [22] D. A. Lowe and L. Thorlacius, *Phys. Rev.* **D60** (1999) 104012, [hep-th/9903237](#).
- [23] L. Susskind, [hep-th/9901079](#); J. Polchinski, [hep-th/9901076](#).
- [24] R. Bousso, *Rev. Mod. Phys.* **74** (2002) 825–874, [hep-th/0203101](#).
- [25] R. M. Wald, *Living Rev. Rel.* **4** (2001) 6, [gr-qc/9912119](#).
- [26] R. M. Wald, “General relativity,” Chicago Univ. Press (1984) 491p.
- [27] E. E. Flanagan, D. Marolf, and R. M. Wald, *Phys. Rev.* **D62** (2000) 084035, [hep-th/9908070](#).
- [28] W. Fischler and L. Susskind, [hep-th/9806039](#).
- [29] L. Susskind and E. Witten, [hep-th/9805114](#).
- [30] See *e.g.* J. McGreevy and H. Verlinde, *JHEP* **12** (2003) 054, [hep-th/0304224](#).
- [31] See *e.g.* N. Goheer, M. Kleban, and L. Susskind, *JHEP* **07** (2003) 056, [hep-th/0212209](#); T. Banks and W. Fischler, [hep-th/0310288](#).